

Visual object recognition is facilitated by temporal community structure

Ehsan Kakaei^{1,2,3,*}, Stepan Aleshin^{2,3} and Jochen Braun^{2,3}

¹ ESF graduate school ABINEP, Otto-von-Guericke University, Leipziger str. 44, Magdeburg 39120, Germany

² Institute of Biology, Otto-von-Guericke University, Leipziger Str. 44, 39120 Magdeburg, Germany

³ Center for Behavioral Brain Sciences, Otto-von-Guericke University, Leipziger str. 44, Magdeburg 39120, Germany

*Corresponding author email (ehsankakaei91@gmail.com)

Supplemental Materials and Methods

Visual similarity of objects

To quantify the visual dissimilarity of the complex 3D objects, first we take 6 snapshots along the 3 principal axes (obtained from PCA). Then, we greyscale the snapshots where the lower values correspond to more depth (Figure S1A). Finally, we calculate the pixelwise Euclidean distance between these grey-scaled snapshots and normalize it by the maximum possible distance. In order to generate distinguishable objects as familiar objects, we discarded any object that had low dissimilarity with the previously generated objects. The threshold used was the 2/3rd quantile of the dissimilarity distribution (computed over >12,000 random pairs). In other words, we accepted new objects only from the top 33% of this distribution.

We assigned familiar objects randomly to each node in the modular graph. Therefore, we didn't expect to observe any significant correlation between object dissimilarity and sequence structure. Indeed, dissimilarity analysis revealed no significant difference either between objects within and between modules (Figure S1B), or between objects pairs at different nodes (Figure S1C).

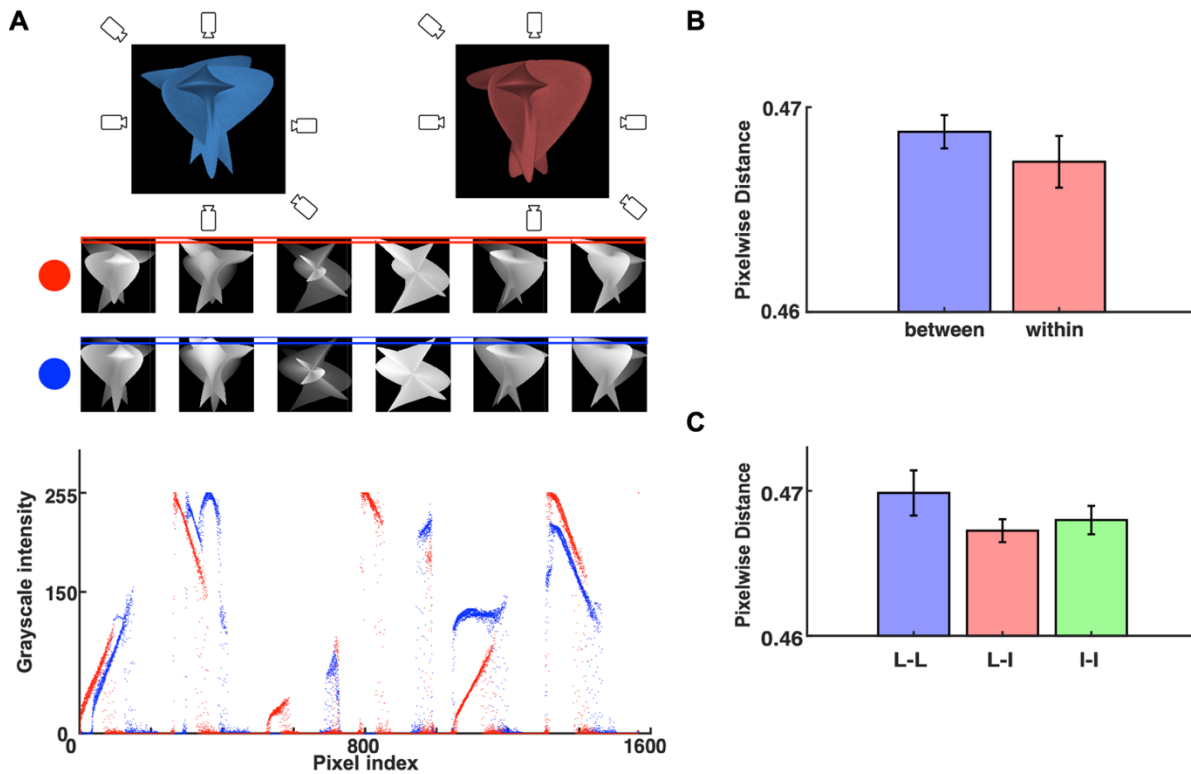


Figure S1 Visual dissimilarity of 3D objects. **(A)** To quantify the visual dissimilarity of two objects 6 grey-scaled snapshots (*middle*) of each object (*top*) are taken in 3 principal directions. Blue and red dots in the graph (*below*) illustrate the greyscale intensity of each pixel for the blue and red objects, respectively. **(B)** The objects within a module are as dissimilar as the object between modules. **(C)** The dissimilarity of objects at linking and internal nodes (L-I) are comparable with to the dissimilarity of objects at linking nodes (L-L) and at internal nodes (I-I).

Object recognition learning task

Experiment 1:

Six observers participated in 8 sessions of recognition learning task in 2 weeks (4 sessions per week). Observers were instructed to indicate their level of familiarity to a novel object by pressing buttons assigned to 'familiar', 'unfamiliar' and 'not sure' responses. Each session was composed of 6 sequences (each with 180 presentations of 2.5 s duration, 450 s total duration). Sequences were separated by a few minutes break. During a given week, participants observed either 'strongly structured' or 'unstructured' presentation sequences. Observers received no training sessions prior to the experiment.

Experiment 2:

Eight observers participated in 6 fMRI scanning sessions while performing the recognition learning task in 2 weeks (3 sessions per week). Observers were instructed to only respond 'familiar' or 'unfamiliar' by button presses. Each session comprised 6 sequences (each with 200 presentations of 3 s duration, total duration 600 s). During a given week, participants observed either 'strongly structured' or 'unstructured' presentation sequences. Observers received one training sessions with sequences of 2D images prior to the experiment (see below).

Experiment 3:

Six observers participated in 6 sessions of recognition learning task in 2 weeks (3 sessions per week). Observers were instructed to only respond ‘familiar’ or ‘unfamiliar’ by button presses. Each session comprised 6 sequences (each with 200 presentations of 3 s duration, total duration 600 s). During a given week, participants observed either ‘strongly structured’ or ‘weakly structured’ presentation sequences. Observers received one training sessions with sequences of 2D images prior to the experiment (see below).

Training session for experiment 2 and 3

Training sessions were similar to the main experiment (6 runs, ‘familiar’ vs. ‘unfamiliar’ responses), except that two-dimensional (2D) objects were used. During each run, observers viewed 120 presentations of the 10 recurring objects and 20 presentations of (20) nonrecurring objects. The presentation sequence was generated as a pseudo-random walk on a fully connected graph (but without direct repetitions). At the end, observers were informed of their total recognition performance (which was near ceiling for all participants).

Response time analysis

We consider the response time (RT) as the time that it takes an observer to press a response button after the onset of the trial, irrespective of the type of the response. In case of no response, we take the duration of the trial as the response time (1 - 7% of responses). Initially, RT did not differ between ‘unstructured’ and ‘strongly structured’ sequences. In the later sessions, responses to the objects were faster in ‘strongly structured’ sequences than the ‘unstructured’ sequences. However, responses to the objects were faster in ‘weakly structured’ sequences than the ‘strongly structured’ sequences (Fig S2).

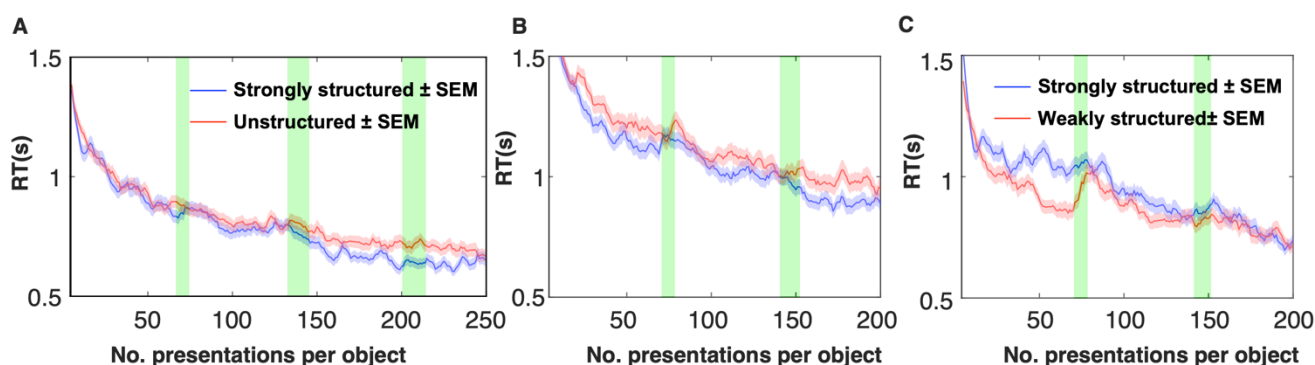


Figure S2 Response time analysis. Response time (RT) is calculated from the onset of each trial. We used a sliding window of width 5, over consecutive presentations of each object. (A) Average RT during experiment 1. We do not observe any significant difference between the RTs during strongly structured and unstructured sequences, over the first 2 sessions (green shades). Starting from the third session, average RT under the strongly structured sequence decrease more than the RT under the unstructured sequence. (B) Average RT during experiment 2. Starting from the first session, RT under the strongly structured sequences are always lower than RT under unstructured sequences. (C) Average RT during experiment 3. Interestingly, the RT under weakly structured sequences were faster than strongly structured sequences.

A

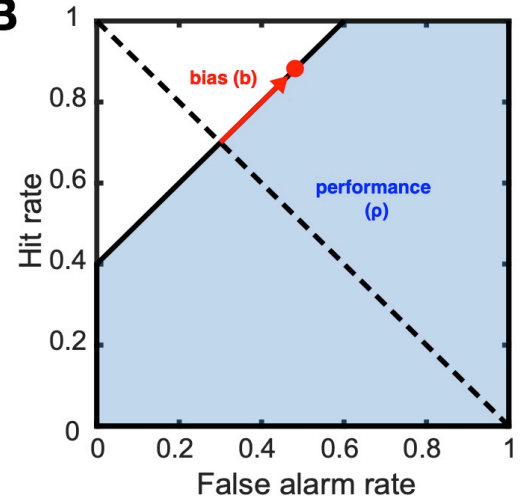
False alarm rate: 0.6

Sequence: ADCABDACEBA....

Response: 0010100101....

r_A : 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1...

$w = 5$

B**C**

r_A	0	0	0	1	0	0	1	0	1	1	1
h_A	--	--	--	--	0.2	0.2	0.4	0.4	0.4	0.6	0.8
ρ_A	--	--	--	--	0.18	0.18	0.32	0.32	0.32	0.5	0.68
b_A	--	--	--	--	-0.14	-0.14	0	0	0	0.14	0.28
H_A	--	--	--	--	0.68	0.68	0.9	0.9	0.9	1	0.9

Figure S3 Object recognition performance analysis. **(A)** To calculate the performance of an observer in recognizing a given object A, we obtain the hit ratio h_A from the number ‘familiar’ responses n_f in N_w consecutive presentations of object A, and compare the false alarm ratio f of the current session. Here, $f = 0.6$ and $N_w = 5$. **(B)** Given point (f, h) , corrected performance $\rho_A = 1 - (f + (1 - h))^2 / 2$ is computed as the area beneath a diagonal line through (f, h) and bias $b_A = (f + h - 1) / \sqrt{2}$ is computed as the distance of (f, h) to the off-diagonal. **(C)** The table shows a representative sequence of familiarity responses and the resulting values of hit ratio h_A , performance ρ_A , bias b_A , and performance entropy H_A .

Object recognition performance

To computed observer performance in recognizing a familiar object, we used a simplified sensitivity analysis (Macmillan and Creelman 2004; Figure S3). False alarm ratio $f = 1 - \frac{n_u}{N_u}$ was obtained from the number of correct rejections ‘ n_u ’ (presses of ‘unfamiliar’ button) and the total number of unfamiliar objects ‘ N_u ’ presented during the current session. For each familiar object, a hit ratio $h = \frac{n_f}{N_w}$ was computed from the number of correct recognitions ‘ n_f ’ (presses of ‘familiar’ button) within sliding windows with $N_w=5$ consecutive presentations of the object in question. Note that, between two consecutive presentations, numerous other objects were presented. Corrected performance was computed as probability correct $p = 1 - \frac{(f + (1 - h))^2}{2}$ and bias as $b = \frac{(f + h - 1)}{\sqrt{2}}$, as illustrated in Figure S3B. Note that the hit ratio often equaled unity ($n_f = N_w = 5$), making computation of z-scores and d' values impractical. Alternative ways of working around this problem and permitting the computation of performance measures A' and d' did not materially alter the results (Stanislaw and Todorov 1999).

To identify the point in time at which an observer starts to recognize a particular object as familiar ('onset of familiarity'), we analyzed sliding windows of $N_w = 5$ consecutive presentations of the object in question. A 'low threshold' definition of 'onset' was the window with peak performance entropy $H_p = -[\rho \log_2(\rho) + (1 - \rho) \log_2(1 - \rho)]$ or, equivalently, the window in which corrected performance ρ transitioned from near chance to near unity. An alternative, 'high threshold' definition was the window in which performance first surpassed $\rho = 0.875$. This threshold corresponds to 100% hits (in $N_w = 5$ presentations) and <50% false alarms (in $N_u = 54$ presentations). Other threshold choices did not materially change the results.

Order of 'onsets of familiarity'

To analyze the ordering of 'onsets of familiarity' (defined by either high- or low-threshold approach), a presentation sequence of recurring objects during one week (3 or 4 sessions) was described by value triplets with object ID/graph node ID (1 to 15), community ID (1 to 3), and onset flag (0 or 1). To obtain the rank order of all 'onsets of familiarity', onset presentations were extracted from the full sequence (and all non-onset presentations were discarded).

Next, pairs of successive 'onsets' were compared based on graph node ID (adjacent or non-adjacent objects?) and community ID (same or different community?) and the probability of successive onsets in each category (same or different community, adjacent or non-adjacent on graph) was obtained by combining data from the presentation sequences of all observers in one condition.

To assess significance, we shuffled the order of the value triples representing each 'onset of familiarity'. Note that this procedure maintains graph geometry (community ID and graph node ID). The p-values for different subcategories (same/different, adjacent/non-adjacent) were corrected for false discovery rate (FDR, Benjamini and Hochberg, 1995).

In an alternative analysis, we established the distribution of intervals between successive onsets (number of presentations separating two onsets) from the presentation sequence of all observers in one condition. We compared this distribution to a null distribution obtained by shuffling presentation order 1000 times (for every observer in one condition). The comparison showed that shorter intervals between onsets occurred significantly more frequently in highly structured sequences (with temporal communities) than in shuffled sequences. No such difference was observed for either weakly structured or unstructured sequences.

Validation task

The validation task comprised 60 trials. In each trial, observers viewed an array of 12 different (rotating) objects of the color corresponding to the current condition (red or blue). Three of these objects were recurring (i.e., had been presented at least 250 times during the preceding sessions), whereas the other nine were novel (never seen before).

Observers were instructed to indicate (by mouse click) the 3 objects that they considered most familiar. Observers were given 30s for this task and objects changed color (to dark yellow) at the moment of selection. After three objects had been selected, observers received binary auditory and visual feedbacks (all objects correct, some or all objects incorrect). Each of the 15 recurring objects appeared in exactly 12 trials.

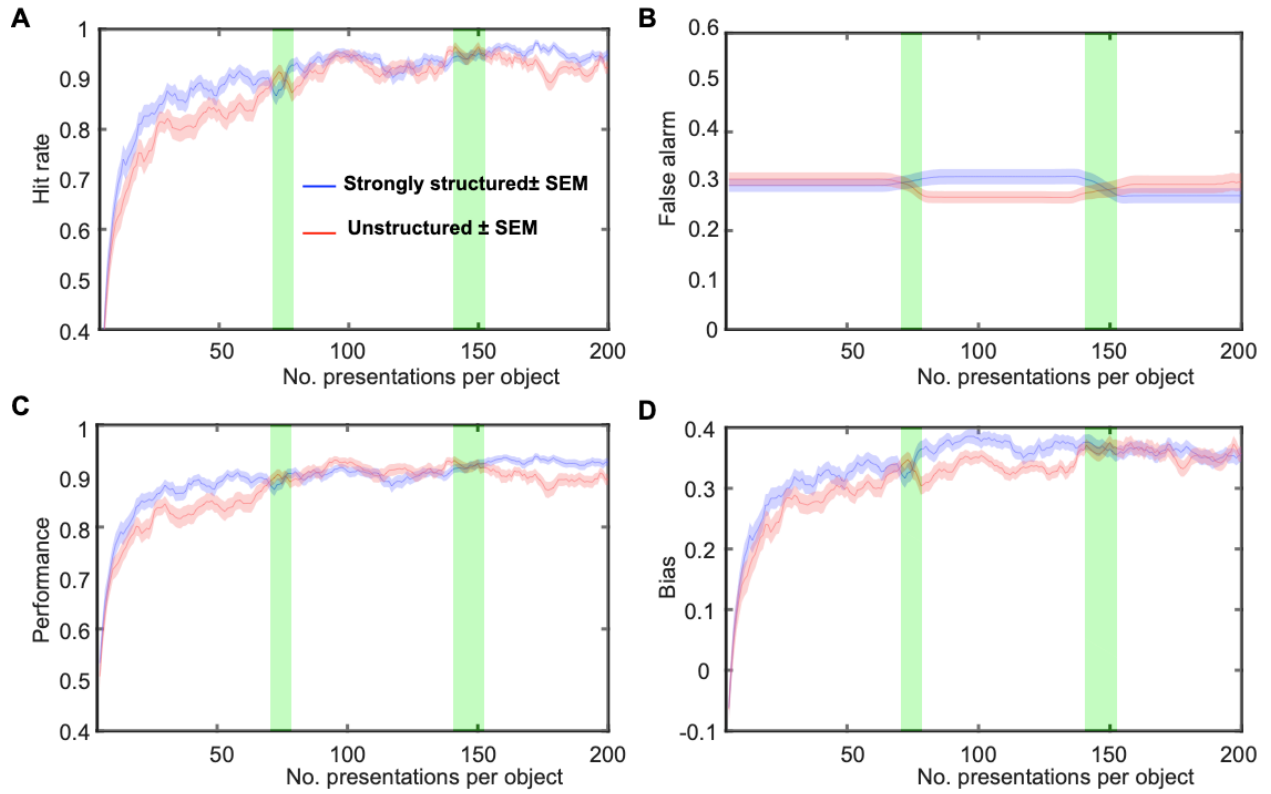


Figure S4 Object recognition performance of observers during experiment 2. Average proportion of hits (recurring categorized as familiar, per window) increases by higher number of presentations per object, while (B) Average proportion of false alarms (non-recurring NOT categorized as unfamiliar, per session) stays almost constant. (C) Average corrected performance p increases nearly monotonically with the number of presentations of a given object. The difference between conditions becomes significant ($p < 0.05$) after approximately 30 presentations. (D) Average criterion bias b , as a function of presentation number. Observers adopted more liberal criterion during later sessions. Green regions indicate the transition between sessions (from last trial of 20% to 80% of all objects in previous sessions).

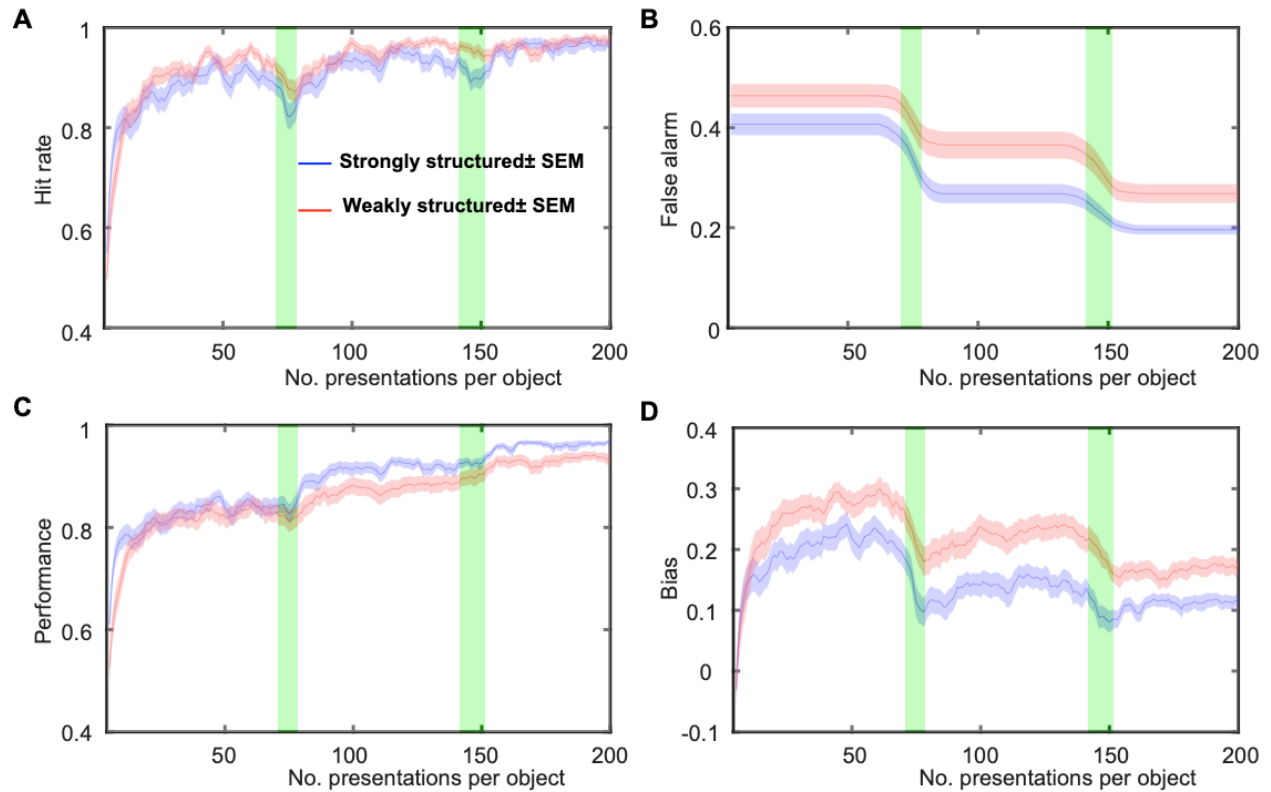


Figure S5 Object recognition performance of observers during experiment 3. (A) Average proportion of hits (recurring categorized as familiar, per window) increases with the number of presentations of a given object. (B) Average proportion of false alarms (non-recurring categorized as NOT unfamiliar, per session) decreases with the number of presentations. (C) Average corrected performance ρ increases nearly monotonically with presentation number. Its value was consistently superior for strongly structured sequences (with temporal community structure) compared to weakly correlated sequences. The difference became significant ($p < 0.05$) after approximately 70 presentations. (D) Average criterion bias b , as a function of presentation number. Observers rapidly adopt a liberal criterion, which abates somewhat during later sessions. Green indicate the transition between sessions (20% to 80% of objects in previous session).

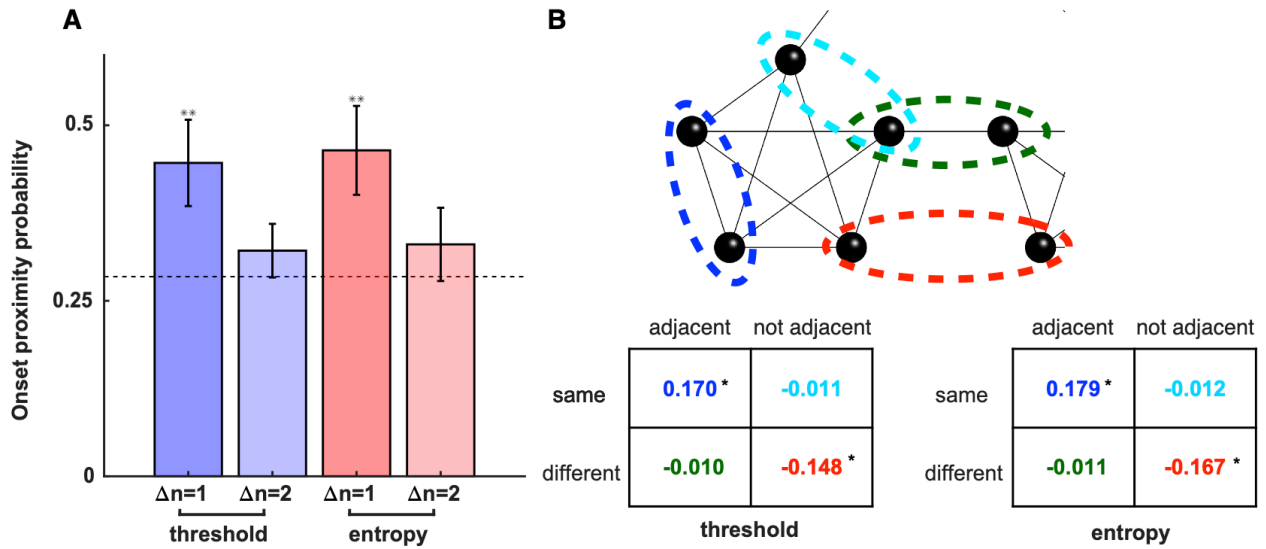


Figure S6 Analysis of the onset of familiarity with individual objects in Exp. 2. (A) Successive onsets of familiarity ($\Delta n = 1$) are far more likely (** $p < 0.005$) for objects within the same cluster than would be expected by chance (dashed line). (B) Frequency of successive onsets, relative to chance level. Object pairs may be in the same or in different clusters, and may be adjacent or non-adjacent on the graph. Frequency is significantly elevated (* $p < 0.05$, after FDR correction) for adjacent objects in the same cluster (blue), but suppressed for non-adjacent objects in different clusters (red).

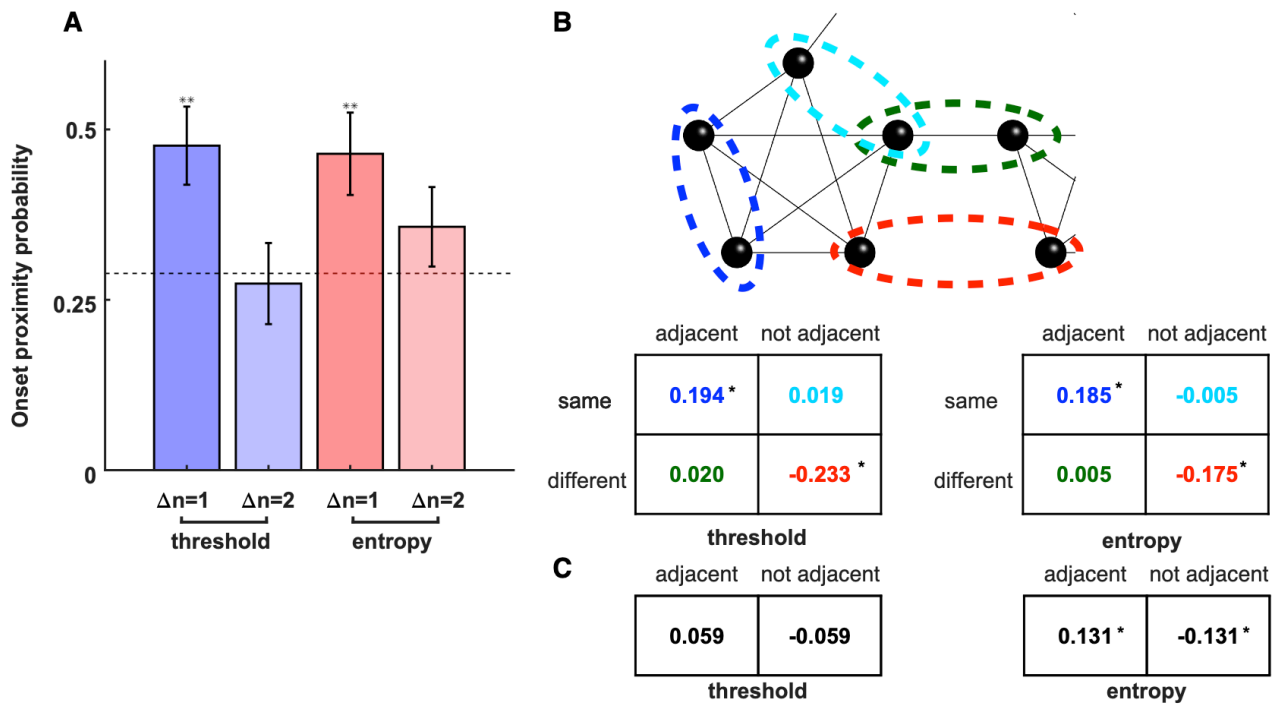


Figure S7 Analysis of the onset of familiarity with individual objects in Exp. 3. (A) Successive onsets of familiarity ($\Delta n = 1$) are far more likely (** $p < 0.005$) for objects within the same cluster than would be expected by chance (dashed line). (B) Frequency of successive onsets, relative to chance level, in *strongly structured* sequences. Frequency is significantly elevated (* $p < 0.05$, after FDR correction) for adjacent objects in the same cluster (blue), but suppressed for non-adjacent objects in different clusters (red). (C) Frequency of successive onsets, relative to chance level, in *weakly structured* sequences. Object pairs may be adjacent or non-adjacent on the graph. Frequency is significantly elevated (* $p < 0.05$, after FDR correction) for adjacent objects, if onsets are entropy-defined.

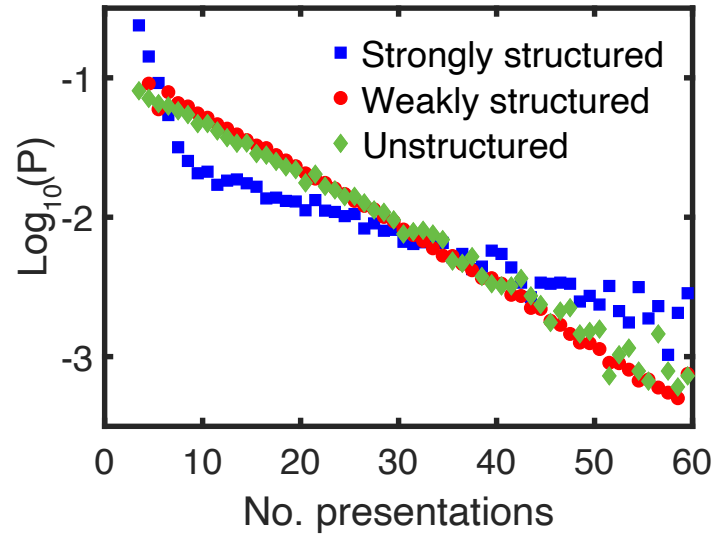


Figure S8 Distribution of repetition latency (number of presentations intervening between repetitions) for strongly structured (blue), weakly structured (red) and unstructured sequences (green). In strongly structured sequences, both very early repetitions and very late repetitions are more frequent than in weakly structured or unstructured sequences. Interestingly, the distributions of weakly structured and unstructured sequences are almost identical.